DOCUMENT RESUME

ED 119 605 95 IR 002 744

AUTHOR Clark, Richard E.; Snow, Richard E.
TITLE Alternative Designs for Instructional Technology Research.
INSTITUTION Association for Educational Communications and Technology, Washington, D.C.; Stanford Univ., Calif. ERIC Clearinghouse on Information Resources.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE 76
NOTE 23p.
JOURNAL CIT Audiovisual Communication Review; v23 n4 Winter 1975 p373-394

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage
DESCRIPTORS Educational Research; Instructional Design; *Instructional Technology; *Multimedia Instruction; Research Criteria; *Research Design; *Research Problems; Research Tools

ABSTRACT

Though models of quality research in education have abounded in the past decade, educational technologists have consistently made the same mistakes in attempting to measure the effects of multimedia approaches to teaching: 1) using a treatment that is not a valid implmentation of theory; 2) inadequate observation; and 3) disguising weak findings in the clothes of strong rhetoric. Typical research designs are one short test with skewed samples and no control group, and they usually concentrate on an isolated internal relationship but do not represent the larger reality of the situation. Several models and combination of models, when used accurately, can remedy the deficiencies of this research. (EMH)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED119605

ERIC/AVCR Annual Review Paper[1]

Alternative Designs for Instructional Technology Research

RICHARD E. CLARK

RICHARD E. SNOW

An experimental design is a plan for conducting an experiment. The plan includes: deciding how to assign experimental *units* (e.g., persons, classrooms, schools, districts, etc.) to *treatments* (e.g., to alternative types of instruction or to experimental and control conditions); describing treatments; and deciding what measures to apply to the behavior of the units to assess their responses to the treatment(s). Our ability to draw valid and useful information from experiments rests on the care and insight employed during the

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

Richard Clark, Richard Snow - AVCR

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

[1]This article is the fourth ERIC/AVCR Annual Review Paper, preparation of which was supported by the ERIC (Educational Resources Information Center) Clearinghouse on Information Resources, Stanford University, Stanford, California. The material in this article was prepared pursuant to a contract with the National Institute of Education, U.S. Department of Health, Education and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Points of view or opinions do not necessarily represent the official view or opinion of the NIE.

Richard E. Clark is associate professor of education and chairman of the Area of Instructional Technology, Syracuse University, Syracuse, NY 13210. Richard E. Snow is professor of education and psychology, School of Education, Stanford University, Stanford, CA 94305.

IR 002 744

design stage. Yet, the number and diversity of questions confronted by a field of inquiry tends to increase faster than the research designs available for tackling these problems (c.f., Campbell & Stanley, 1966).

The purpose of this article is to survey the designs currently used in research on instructional technology and then to suggest alternative plans for experiments that may be more effective in attacking the special problems confronted in this field. This is not a comprehensive treatment of research design in the Fisher (1935) or Campbell and Stanley (1966) tradition. The range of problems and alternatives discussed must necessarily be tailored to limited space, with citations to more complete discussions for those who wish to pursue particular issues. The article is frankly aimed at influencing those who teach research design in graduate schools. At the same time, it is hoped that the approaches described here will interest instructional technologists who wish to expand the range of their problem solving techniques.

A SURVEY OF AVCR RESEARCH DESIGNS

To determine the variety of current designs employed in instructional technology experiments, a descriptive survey of the last five years of *AV Communication Review* (AVCR) was conducted. Each article reporting data was categorized according to: 1) type of study; 2) design employed; 3) number of units or subjects surveyed in each study; 4) amount of time spent with subjects; and 5) completeness of treatment descriptions.

Of the 111 articles reviewed, 49 (44 percent) were studies that reported data collected in original experiments and 62 (56 percent) were concerned with theory, literature reviewing, discussion, etc. Table 1 displays the original authors' descriptions of the type of article being presented. In most cases it was relatively easy to place each one of the

TABLE 1
AVCR Authors' Description of Article Content

Description	*Number of Articles*	*Percent of Total*
A. Experimental Study	25	23
B. Evaluation	13	12
C. Correlational	6	5
D. Theory, position paper, etc.	62	56
E. Mixed (A/B; A/C; B/C)	5	4
Total	111	100

TABLE 2
Types of Designs Used in AVCR Studies (Following Campbell & Stanley, 1966)

	Design Type[a]	Number of Studies	Percent of Total
I. Pre-experimental designs			
One shot	X O	15	31
One group pretest-posttest	O X O	12	25
Static group	$\begin{pmatrix} X & O \\ & O \end{pmatrix}$	6	12
II. True experimental designs			
Pretest-posttest Control group	$\begin{pmatrix} R & O & X & O \\ R & O & & O \end{pmatrix}$	8	16
Posttest only Control group	$\begin{pmatrix} R & & X & O \\ R & & & O \end{pmatrix}$	2	4
III. Quasi-experimental designs			
Counterbalanced	$X_1O \quad X_2O$ $X_2O \quad X_1O$	2	4
IV. Correlational designs		2	4
V. Unable to Categorize		2	4
Total		49	100

[a]Notation for design type is as follows: X = treatment, O = pretest or posttest, R = random assignment (after Campbell & Stanley, 1966).

experimental studies in one of the design categories suggested by Campbell and Stanley (1966). Table 2 presents the categorization for the 49 studies of concern here.

The survey identified several shortcomings in current practice. These are discussed in the sections that follow.

Description, Duration, and Sample Size Reporting

A critical problem in instructional technology research is the specification of treatments. The reader needs to know a) what the treatment entailed, and b) whether the treatment was a valid operationalization of theoretical constructs. Table 3 indicates our judgment of the adequacy of treatment descriptions in each study.

TABLE 3
Treatment Descriptions in AVCR Studies from 1970-1975

Completeness of Description	Number of Studies	Percent
1. Full specification—possible to replicate	8	16
2. Some additional information required	13	26
3. May be replicable with added information	12	25
4. Probably not replicable	10	21
5. Obviously not replicable	6	12

Another design element important in judging the value of experimental information is the amount of time students spend in a treatment. This is particularly critical when the research concerns methods of designing or presenting instruction in the classroom. In the 49 AVCR studies surveyed, treatments appeared to range from three to 2000 minutes in length, with an average of 95 minutes. However, four exceptionally long studies accounted for this average. When these were removed, the mean reduced to 24.7 minutes for 45 studies. This is hardly a treatment duration from which one might expect to generalize to *courses* of instruction.

Surprisingly, a relatively large number of students (mean = 126) were used per experiment. Typical experiments in other fields of instructional research often use much smaller samples.

Reporting

No research design ever succeeds in eliminating all threats to validity. Therefore, the investigator usually must decide which potential types of error he or she is willing to tolerate. We are left, then, with imperfect data that contain *anticipated* error and with the ethical responsibility to make those limitations explicit in research reports. In many of the AVCR articles that reported evaluation studies, authors seemed to slip easily into prose usually reserved for conclusion-oriented designs. Although in most cases the designs were clearly described, less sophisticated readers might easily be misled by the interpretations.

Asher and Vockell (1973), for example, found that a sample of educational decision makers tended to overestimate the usefulness and quality of research reports that contained serious design problems when authors did not describe possible sources of systematic error. They also found that a sample of researchers tended to give significantly lower quality ratings than did the decision makers to the same studies. It appeared that the situation resulted from authors' attempts to give some vitality to their reports rather than from any willful deception. Guarded language and constantly qualified interpretations do not make research reports terribly exciting reading to the uninitiated. Nor does this form allow the author much chance to relate hunches, hypotheses, or feelings about what actually went on (or what might have gone on) in the experiment.

Yet, one cannot simply "let the consumer beware." A balanced solution may be to assign separate spaces for the "consumer report" and the "author's corner." The first gives a solid and conservative report of the detailed data and findings, including full discussion of threats to internal and external validity, events in the collection of data that may have affected outcomes, etc. Here the author acts as his or her own best critic. The second discussion then can be devoted to the author's personal impressions of what happened in the study and how the data might be suggestive for instructional practice or future research. While consumers still may not read both sections, at least all the necessary and potentially important information is presented. And no one is prevented from forming his or her own conclusion.

Research Designs

The most common design problem in a majority of AVCR studies was the reliance on pre-experimental plans, i.e., one-shot case studies, static group comparisons, etc. In all of the 33 studies using these designs, none included random assignment of subjects (or units such as classrooms) and few used control groups for comparisons. In some instances, two treatments were compared without random assignment of subjects or control groups. In a few studies this problem was complicated by nonrandom subject (or unit) attrition between pretest and treatment or during successive treatment application and posttest.

Filep and Schramm (1970), in a survey of research studies funded by NDEA Title VII funds between 1958 and 1968, found much the same pattern: "there was a predominant dependence on accidental selection of the sampling unit [p. 97]." In addition, most of the NDEA media studies were of the pre-experimental variety. Over the ten-year period Filep and Schramm studied, they noticed a trend toward moving from field to laboratory experiments (we noticed no such trend in the AVCR sample) but detected no associated improvement in the number or quality of research designs used (also reported in Hall, 1972). One is tempted to conclude that the variety of designs available to instructional technologists has not changed appreciably in the past 15 to 18 years, despite the many advances accomplished in closely related fields. The responsibility for this "steady state" would appear to rest in the university programs where researchers are trained.

Since the sources of invalidity in pre-experimental de-

signs have been discussed by a number of authors (c.f. Issac & Michael, 1972; Winer, 1971) we need not attempt a complete description of the problems of interpretation they pose. The pre-experimental strategy is exploratory in nature, not confirmatory, and should probably not be reported in isolation except under extraordinary circumstances. The results of such studies are difficult, often impossible, to interpret. While they may provide important sources of descriptive data for evaluation purposes, researchers interested in generalizations should consider alternatives to formal experimental designs only when control procedures such as random assignment are *impossible* to accomplish. In such instances, many quasi-experimental arrangements are available (c.f. Blalock, 1964; Baker & Schutz, 1972; Campbell & Stanley, 1966; Issac & Michael, 1972; Riecken & Boruch, 1974). Although quasi-experiments are preferable to pre-experimental or ex post facto techniques, they should be considered only when all possible routes to fully controlled experimentation have been explored and rejected as either too expensive, impossible because of logistical problems, or unrepresentative of the environment in which the treatment is to be used. In most instances, problems of finances and logistics might better be met by seeking institutional or governmental support rather than by design compromises. Concern about representativeness or generalizability, however, is a more justifiable basis for considering alternatives to formal experiments. It is to this issue that the discussion next turns.

Internal vs. External Validity

It often appears when choosing between laboratory or field research settings, and the designs that seem possible in each, that we must opt for either internally *or* externally valid plans; i.e., that one can't have both. And the traditional view has usually argued for securing internal validity at the expense of external validity—for systematic control at the expense of representativeness. Pereboom (1971) is among those who have criticized this orientation, noting that "if complex behavior is assumed to be both probabilistic and multidimensional, 'stripping' the environment down to a minimum in order to control, to determine the role of a few variables, may be a potentially self-defeating process [p. 445]." Cronbach (1975) and Ebel (1967) among others also seem to conclude that the search for generalizable conclusions based on analy-

tic research is futile. These critics often go on to suggest one or another class of quasi-experimental designs or to emphasize descriptive or decision-oriented evaluation studies.

For many, however, systematic analysis, control, and internal validity remain the *sine qua non* of research.

But it may be possible to avoid the dilemma by constructing research designs that attempt to obtain internal validity without sacrificing representativeness. In the past few years an increasing number of discussions and design suggestions have sought to incorporate both concerns (e.g., Bracht & Glass, 1968; Campbell, 1969; Shulman, 1970; Snow, 1974; Baker & Schutz, 1972; Buss, 1974; H. Clark, 1973; R. Clark, 1975; and Salomon & Clark, 1976). In the next section we discuss abbreviated versions of those designs that appear to be most useful to instructional technology researchers.

COMBINED DESIGNS IN FIELD SETTINGS

Natural settings provide researchers with a host of situations, outstanding events, and highly innovative projects, which deserve to be carefully studied.[2] Such possibilities are particularly prominent in the domain of instructional technology. Their investigation is important inasmuch as their quality, imaginativeness, and complexity far exceed the events typically studied by researchers. Most investigations of innovative, yet complex, real-life instructional materials or techniques, if conducted at all, are usually limited to relatively simple and gross evaluation studies that often lack internal validity. The problem is therefore how to conduct research on real-world events, including large-scale program evaluations as well as research into the effects of outstanding innovations, while attaining satisfactory internal validity.

The Concomitant Variation Design

Salomon (1971) suggested that media research might sometimes be more successful if it started out with events in the real world and worked *backwards* into the laboratory by gradually analyzing them into ever smaller components. Earlier, Shulman (1970) proposed a similar approach labeled the "Epidemiological Strategy" to accomplish the same end for research in teaching and learning. Essentially, outcomes are compared after learners have been differentially exposed to an external, natural, factor such as a TV program.

[2]Many of the concepts in this section were taken from an unpublished manuscript by Salomon and Clark (1976). The authors wish to acknowledge their debt to Gavriel Salomon of The Hebrew University, Jerusalem, Israel.

Gathering data on numerous individual difference variables including personal background, abilities, prior achievements, and the like, it should be possible to distinguish between those who are more and those who are less affected by the program.

This strategy needs to be supplemented by a careful analysis of the various components of the program or technology. Identifying such significant components, the researcher should be able to generate hypotheses as to their possible effects and effectiveness. We should thus be able to investigate not only *who* was more and who was less affected but also *what caused* the effect. In this way, a real life event could be studied *as if* carefully controlled experimental conditions were present, while in fact they were not.

The Concomitant Variation Design (somewhat different from Shulman's Epidemiological Strategy) is based on the measurement of three kinds of independent variables: relevant individual differences of the students involved, instructionally significant components of the program, and amount of student exposure to, or involvement in, the program. Students differ as to the amount of their exposure to, or involvement in, an instructional program. Thus, exposure is a continuous, major independent variable. The purpose of this approach is, then, to examine the extent to which amount of exposure or involvement differentially affects students. Since, however, the program is analyzed into its significant components, one can also address the question of what elements in the program affect individual learners.

It is clear that the examination of the program's effects, when carried out under natural conditions, is methodologically deficient. The amount of exposure to the program by each student may be the result of self-selection. More able, curious, or motivated students may choose to expose themselves more to the program. The necessary condition of "other things equal" is not met but statistical procedures may be used to examine the most likely self-selection hypotheses. Toward this end, background data need to be collected and multiple-regression procedures used (c.f. Cohen, 1968). It then becomes possible to *partial out* initial exposure-related differences. We approximate the condition of "other things being equal" through the Concomitant Variation procedure rather than through other more familiar design procedures.

An Example. The introduction of *Sesame Street* to Israeli children created a unique opportunity to study the effects of a highly complex and sophisticated program on children who were naive with respect to TV (Salomon, 1974). Since, however, the program was broadcast simultaneously all over the country, a traditional experimental design was impossible. No adequate control group of children who were not expected to watch the program could be formed. On the other hand, simple comparisons between heavy and light viewers of the program would be meaningless, since the amount of viewing could be the result of self-selection.

Even if this threat to internal validity was removed, there was still the problem of external validity. Since the effects of only one program were to be studied, generalizability would be limited, as in most evaluation studies. The effects of one program might not represent the possible effects of other programs among a total of 40 one-hour shows.

Some statistical methods, however, allow us to reduce these difficulties. To study changes in achievement presumed to be the result of program viewing, each child's amount of exposure to the program was measured and the degree to which exposure related to later achievement was computed. Here is a situation where the independent variable (exposure) has values distributed over a wide range, from total non-exposure, through many levels of partial exposure, to total exposure to every show. In this respect we have an advantage over the traditional experimental procedure in which children might be divided into groups of "viewers and nonviewers." The traditional approach usually ignores differences *within* each one of the groups, whereas here they are taken into account.

The statistical method of multiple regression allows us to partial out the contributions of background and initial achievement variables, thus measuring the "net" contribution of exposure to the post-viewing achievements (e.g., Cohen, 1968). In other words, we are able to specify the "net" amount of post-viewing achievements which can be attributed to exposure, other things being equal—to an extent. If premeasures have been selected with care, at least the major self-selection hypotheses can be accounted for.

This method of analysis also allows us to compare groups

since the same background and pre-viewing measures can be entered into the analyses in the same order, and exposure entered as the last predictor. It is then possible to see in which group its "net" contribution to post-viewing achievement is larger.

Table 4 provides an abbreviated example of such an analysis for data on two criterion tests of program effects. As can be seen, all background and initial achievement measures accounted for 36.8 to 50.8 percent of the post-viewing variance, depending on the group and the test. Exposure accounted for an *additional* 4.3 to 16.3 percent. It is also seen that while exposure made a significant difference for lower class children in the case of the Letter Matching test, it did not make much of a difference for middle class children. The converse is true in the case of the Parts of the Whole test.

The question of generalizability was addressed through conceptualization of specific program components, followed by the generation of specific hypotheses. Thus, for example, it was hypothesized that particular presentation formats used in the program would affect specific skills in particular children. Although such components could not be experimentally manipulated, it was still possible to test such hypotheses using multiple regression procedures. Of course, this procedure should be regarded as *exploratory* in nature. It is useful for deriving hypotheses for further study and for making decisions about the worthwhileness of broadcast programs.

TABLE 4
Amount of Post-Viewing Variance Accounted for by Background, Initial Achievement, and Exposure (After Salomon, 1974)

Variance Accounted for on Test of . . .	*Source of Variance*	*All Background Variables* R^2	*All Previewing Tests* $+R^2$	*Total* R^2	*Exposure* $\div R^2$	F
Letter matching	Lower class	26.7%	21.1%	47.8%	16.3%	5.40*
	Middle Class	14.8	36.0	50.8	4.3	1.96
Parts of the whole	Lower class	20.9	27.6	48.5	6.6	3.60
	Middle class	10.0	17.8	36.8	18.3	6.90*

*$p < .05$

It is not a formal experimental approach and many sources of invalidity remain uncontrolled.

The Staged Innovation Design

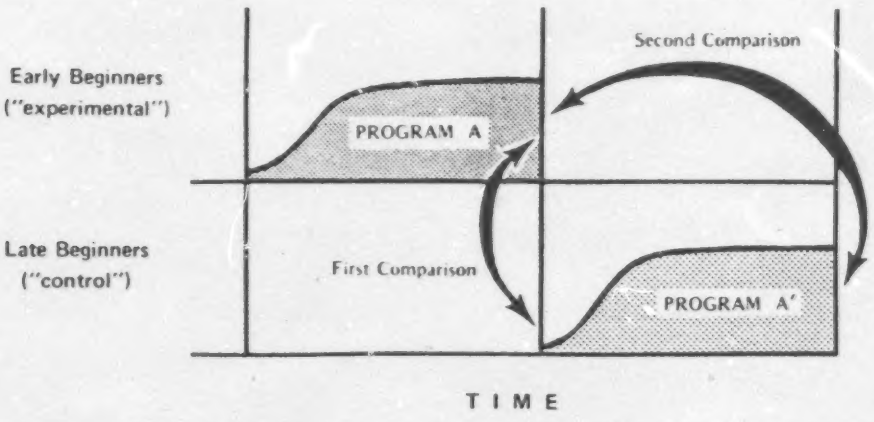
Sometimes it may happen that a new media-based program, sufficiently innovative to deserve a thorough study, is introduced into schools. Let us assume that all students are to participate in the program, again making it difficult to create adequate control groups. However, it might be possible to introduce the program in stages, thus allowing for a *Staged Innovation Design* (Campbell, 1969). Following this design not all schools are introduced to the program simultaneously. Some schools, chosen randomly if possible, are introduced to the program earlier than others. The early beginners turn out to serve as the "experimental" group while the late beginners serve *temporarily* as the no-treatment "controls." "Experimentals" (early beginners) can then be compared with the "controls" (late beginners) on achievement or any other dependent variables.

This design can be further developed as follows: Once the "controls" take part in the program, their achievement can be compared with those of the "experimentals" as measured on an earlier date. Thus, a replication is built into the design. Two groups have taken part in the program, one after the other, and their results can be compared on two occasions: before the "control" schools started out with the program, and again—after they have finished it.

One can also try to change the program before the "control" schools begin to participate in it. Comparing their post-participation results with those of the "experimental" schools, measured on an earlier date, is similar to an experimental comparison in which the newly introduced changes in the program serve as the "treatment." The format for the Staged Innovation Design is shown graphically in Figure 1.

An Example. Elements of the Staged Innovation Design can be found in the Age Cohort Study, part of the first year's evaluation of *Sesame Street* (Ball & Bogatz, 1970). In that study, 114 children, 53-58 months old, were pretested before the program was shown and their achievements compared with those of another group of 101 children of the same age, *after* the program was shown. When the posttest group was divided into viewing quartiles, it was found that those who viewed the program achieved more than the pre-

FIGURE 1
The Staged Innovation Design (After Campbell, 1969)



test group. Thus, the conclusion was reached that viewing the program led to greater gains in scores.

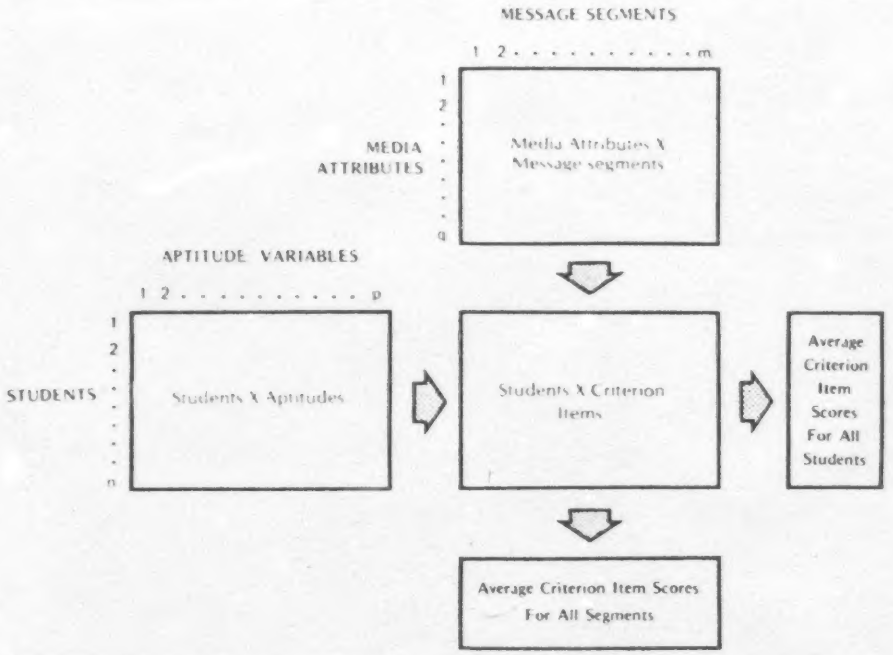
It will be noted that the Age Cohort Study Design resembles the Staged Innovation Design inasmuch as it compares groups at different points in time. Those who are about to receive the "treatment" serve as the controls, and their pre-"treatment" scores are compared with the scores of another group after it has received the "treatment."

Ecological Design

Another design that uses multiple regression techniques to explore viewer and message dimensions simultaneously was first proposed by Seibert and Snow (see Snow, 1974). It uses student background factors as in the Concomitant Variation Design, but also attempts to treat factors varying across segments of the problem in the same way. Potentially, it could be used to take into account any ecological factor that varies across program segments, programs, or other instructional occasions. Hence, a provisional name for it might be the Student-Ecological Interaction Design, or Ecological Design for short.

Background factors including ability, prior achievement, personality, etc. are measured for all students, as before. The program is divided into convenient segments in such a way that each segment can be connected to its own special criterion test or to special items in an overall criterion test.

FIGURE 2
The Ecological Design (Based on Snow, 1974)


MESSAGE SEGMENTS
1 2 . . . . . m
MEDIA ATTRIBUTES
1 2 . . . . . q
Media Attributes X Message segments
APTITUDE VARIABLES
1 2 . . . . . p
STUDENTS
1 2 . . . . . n
Students X Aptitudes
Students X Criterion Items
Average Criterion Item Scores For All Students
Average Criterion Item Scores For All Segments

There are then three data matrixes as shown in Figure 2. Using multiple regression methods, the student aptitudes can be used to predict average achievement scores for each student and the message attributes can be used to predict the average criterion item scores for each segment. This identifies which aptitudes and which media attributes are significant as main effects. Then the student × criterion item matrix can be residualized with respect to these two main effects. In other words, the main effects of aptitudes and message attributes are partialed out of the criterion matrix so that the scores remaining within reflect only performance that is not predictable from aptitudes and message attributes alone. Then, combinations of aptitudes and message attributes can be formed to represent interactions between these two domains, and these combination variables can be used to predict the residualized scores remaining in the criterion matrix.

This is one of a class of experimental arrangements de-

signed to improve the representativeness of research studies. It is built up as a multivariate analog of the basic representative design proposed by Brunswik (1956). A discussion of this point of view is presented by Snow (1974). So far, however, there are no published examples of such designs in instructional technology research, though a dissertation by Heckman (1967) did apply it to an analysis of a programed text.

The Rotation Design

A somewhat simpler version of the Ecological Design may also be appropriate for instructional technology research, where the experimenter wishes to extract some outstanding qualities of media and study their effects and effectiveness in interaction with learners and learning tasks. The researcher generates specific hypotheses concerning the effects of these qualities on particular learners and tries to find out for what kinds of tasks these are most appropriate.

Imagine a program that can be divided into a number of different tasks based on some learning hierarchy, taxonomy, or task analysis. Assume also that a number of general and specific aptitude measures of learners are taken. The researcher then prepares a number of alternative ways of teaching each of the program's tasks such that each task (chapter, topic, or any other discrete component) is taught to another group of learners using a different medium or technology. Each medium prepared to teach the material in one of the task units is so structured as to capitalize on the special attributes of that medium.

Comparable groups of learners, preferably in their natural learning habitats, are taught the same program. However, each group is exposed to different task/medium compositions.

For illustrative purposes we have put into the design four student groups, four media, and a four-task learning program. Other combinations are, of course, possible. Each group learns *the whole program* with a different combination of media. This enables us to compare aspects of learning behavior during the study as well as posttest performance within each row *separately*, that is, within one task and across media. Given that we employ measures common to all tasks (e.g., curiosity), it becomes possible for us to compare results within one medium and across tasks. Finally, aptitude-treatment-interactions can be studied within each

FIGURE 3
The Rotation Design (Task × Medium × Learners Experiment)

LEARNING TASKS	GROUPS			
	G_1	G_2	G_3	G_4
T_1	G_1 A* T_1	G_2 D T_1	G_3 C T_1	G_4 B T_1
T_2	G_1 B T_2	G_2 A T_2	G_3 D T_2	G_4 C T_2
T_3	G_1 C T_3	G_2 B T_3	G_3 A T_3	G_4 D T_3
T_4	G_1 D T_4	G_2 C T_4	G_3 B T_4	G_4 A T_4

* A, B, C, D, = Different media attributes, technologies or different versions of the same medium.

row, thus showing whether learning of a task by means of one medium benefits certain learners in a way different from learning by means of another medium. The same analysis can be carried out within one medium and across tasks and groups (Cronbach & Snow, in press).

In spite of its appearance, this is not a factorial design. Row and column main effects are not of primary interest. For formative evaluation purposes, however, one might study the intercolumn comparison to test the *overall* effectiveness of an instructional package.

Each row in the design represents one learning task, topic, or period of *any* desired duration and complexity. Within the row, a one-way analysis of variance, to test media effects, becomes possible. This could be done with each row separately. However, since learning of the program is cumulative, one might take each row into account in analysis of successive rows using analysis of covariance or multiple regressions. This design could have been used by Allen and Weintraub (1968), for example, to combine their three independent experiments, each dealing with a different task.

An Example. Samuels, Biesbrock, and Terry (1974) wished to determine whether pictorial illustrations would influence beginning readers' attitudes toward stories they read. Some of the psychological effects of illustrations when used in primary readers were investigated earlier, indicating strong interference effects (Samuels, 1970). Thus, the present study was concerned mainly with affective effects and their instructional utility. Using a Graeco-Latin Square Repeated Measures Design, the researchers assigned students to one of three groups. Each group read one story each day for three days. Each story was accompanied by a different type of illustration. Thus, no two groups read the same story with the same illustrations, nor did two groups read the same story on the same day.

The design used by Samuels et al. differed slightly from the Rotation Design, since no particular order of story-presentation was needed. The Rotation Design is better suited to curricula in which chapter or topic order is given. Also, the Graeco-Latin Square design of Samuels et al. does not consider interactions with individual differences. In the Rotation Design this is a critical component. In general, however, the two designs are cut from the same cloth; both permit the study of all possible media/task combinations in natural settings.

The Fractional Design

The large number of subjects required by traditional Fisherian full-factorial designs make them tedious and expensive. Although the use of within subjects designs offsets these problems to some extent, it is impractical to present students with many treatments. This breeds its own problems, through interference among treatments, for example. Fractional designs reduce these problems by allowing the experimenter to use only a fraction of the cells in the full design.

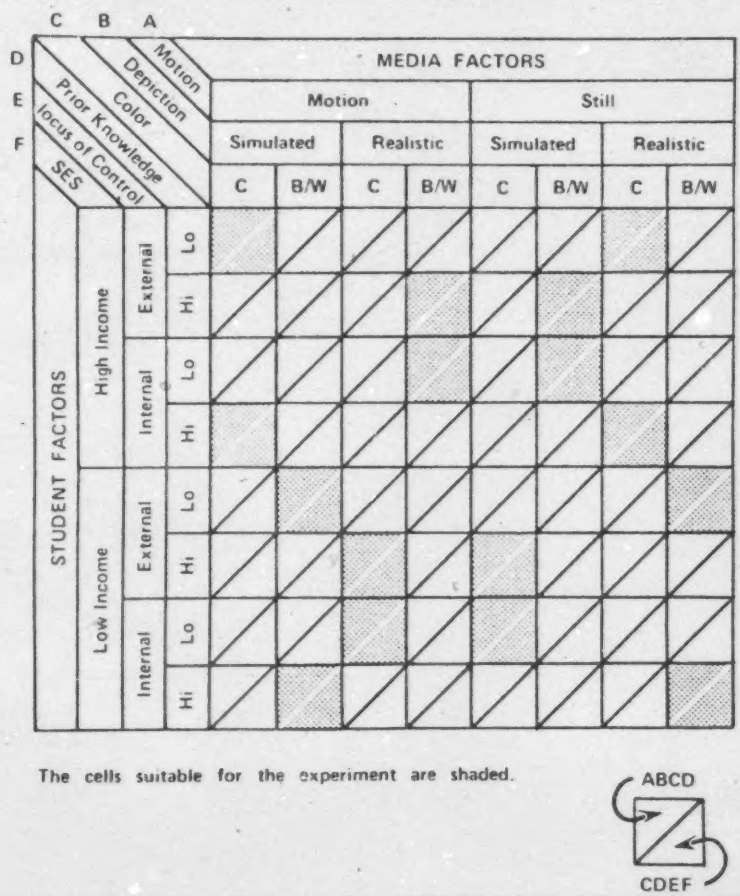
Fractional designs are most useful in pilot research since they allow for the efficient incorporation of many factors that may be of speculative interest or that might inflate the error term unnecessarily if left uncontrolled. For example, a 1/16 fraction of a 2^{10} design:

> . . . allows one to get information on all of the main effects, and first and second-order (two way and three way) interactions with 31 de-

grees of freedom for a pooled estimate of error. This information is obtained from 64 observations instead of the 1024 needed for the full factorial design [Elman, Calfee, & Filby, undated, p. 5].

Calfee (1974) has provided detailed discussion and a number of examples of fractional design in curriculum research and evaluation. He suggests that the approach is particularly useful where theory is "vague, misleading or altogether lacking [p. 13]." Figure 4 depicts one version of this design in a 2×2×2×2×2×2 experiment employing both student and me-

FIGURE 4
A 1/4 Replication of 2^6 Experiment (After Calfee, 1974)



The cells suitable for the experiment are shaded.

dia factors. Even if the experiment were enlarged to include task and/or school variables, the number of cells required still would be considerably less than a full factorial design.

The Intensive Time-Series Design

Another approach, the intensive time-series (ITS) design, may be useful in studying the effects of a treatment over a long period of time in a natural setting where the experimenter has neither abundant resources nor control over school time schedules. This design comes from Campbell and Stanley (1966), as modified by Van Dalen and Meyer (1966) and Thoresen (in press). It was originally constructed to control reactive aspects of pretests. Several measurements of the dependent variable are taken over a period of days (or weeks) before and after the introduction of the treatment. If the series of pretest scores shows no appreciable change across successive observations, one can reasonably assume that treatment effects are not due to maturation, testing, regression changes in instrumentation, or the effects of selection or mortality. The design does not account for the possible interactions between pretesting and the treatment; nor does it control for selection-treatment interactions. The most serious ITS problem, however, is the possible interference of contemporary history. During the time that the treatment is being administered, some event may occur (such as a conversation between subjects related to the treatment, a television program that affects the behavior of the subjects being observed, etc.) that adds to or detracts from the treatment effect. This source of invalidity can be controlled by adding a control subject or group that does not receive the treatment. The control group data also help to examine interactions between selection and maturation.

One of the advantages of this design for instructional technology researchers is its appropriateness for single-subject (or intensive) studies (Thoresen, in press). A teacher, student, administrator, etc. may be observed across a time series in a natural work or study setting, with a minimum of disruption of class and school schedules, at low cost. The design may also be used with groups of subjects or classrooms. It is most appropriately used with treatments that are relatively simple and easy to specify. Those considering the use of this design should consult Thoresen and Elashoff (1973) on appropriate statistical tests.

FIGURE 5
Intensive Time-Series Design (After Alper et al., 1972)

	Teacher Negative Response (Attending to Inappropriate Behavior)		
Baseline (6 days)	Modeling One (5 days)	Feedback One (4 days)	After Phase (6 days)

	Teacher Positive Response (Attending to Appropriate Behavior)		
Baseline (11 days)	Modeling Two (4 days)	Feedback Two (4 days)	After Phase (2 days)

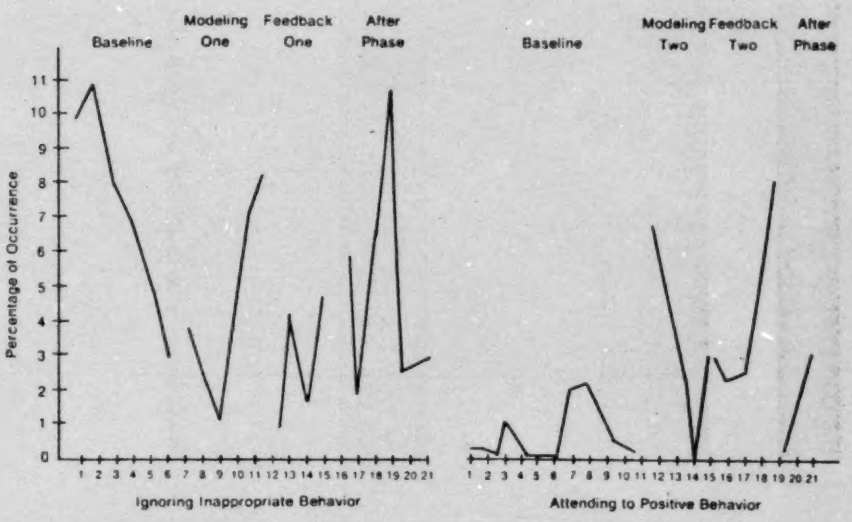
An Example. Alper, Thoresen, and Wright (1972) studied the combined effects of a videotape that modeled ways to increase a teacher's positive attention to "appropriate" student behavior and decrease negative responses to "inappropriate" behavior, and feedback from classroom observers of the teacher's behavior. The design of the experiment is depicted in Figure 5.

The investigators chose the design because previous studies had not been planned in such a way that it was possible to assess the effects of each variable separately on teacher behavior. In addition, the experimenters were curious about the durability of acquired responses over time after training.

Baseline (pretest) data were obtained from one teacher at two points: before training and feedback on ignoring inappropriate behavior and before training and feedback covering attention to positive or appropriate behavior. The modeling tapes were assessed separately from the feedback in both sequences.

As expected, the modeling/feedback treatments produced the desired effect. Unexpectedly, however, when trend analysis (or split level analysis following Thoresen and Elashoff, 1973) was applied, it was found that the desired behavior tended to disappear and the undesired behavior tended to reappear over time unless feedback was used. Other designs often used in such studies would not have been able to uncover this trend. The time-series feature of this study also clearly displayed an unstable trend in the change of teacher behavior over time. This would suggest that there

FIGURE 6
A Display of Time-Series Data (From Alper et al., 1972)



were other factors influencing the observed behavior in addition to the manipulated variables. Figure 6 displays the time-series data from the study.

CHOOSING AND MODIFYING DESIGNS

One of the most important aspects of research is the choice of a specific plan for making those observations presumably related to the theoretical constructs and relationships of interest. Although this article has suggested some designs worth adding to the repertoire of the instructional technology researcher, it cannot foresee the many special considerations and modifications that will be needed in any specific application. No research design should be rigidly or blindly applied. It is up to those who plan studies to modify designs to fit the question being asked. The more the researcher is aware of alternative designs, the better he or she will be at fitting the design to the purpose at hand.

REFERENCES

Allen, W. H., & Weintraub, R. The motion variables in film presentations: Final report. University of Southern California, 1968. (Results also reported in *AV Communication Review*, 1970, *18* (1), 118.)

Alper, T., Thoresen, C. E., & Wright, J. The use of film-mediated modeling and feedback to change a teacher's classroom re-

sponses. R & D Memorandum #91. Stanford University, School of Education, August 1972.

Asher, W., & Vockell, E. Information quality and educational decision making. West Lafayette, Ind.: Purdue Research Foundation, February 1973. ERIC ED 086 214.

Baker, R., & Shutz, R. (Eds.) *Instructional product research.* American Book Company, 1972.

Ball, S., & Bogatz, G. A. *The first year of Sesame Street: An evaluation.* Princeton: Educational Testing Service, 1970.

Blalock, H. M., Jr. *Causal inferences in nonexperimental research.* Chapel Hill: The University of North Carolina Press, 1964.

Bracht, G. M., & Glass, G. V. The external validity of comparative experiments in education and the social sciences. *American Educational Research Journal,* 1968, *5,* 437-474.

Brunswik, E. Perception and the representative design of psychological experiments. Berkeley: The University of California Press, 1956.

Buss, R. A. A general developmental model for interindividual differences, intraindividual differences, and intraindividual changes. *Developmental Psychology,* 1974, *10* (1), 70-78.

Calfee, R. C. The design of experiments and the design of curriculum. Prepublication draft. Stanford University, School of Education, 1974.

Campbell, D. T. Reforms as experiments. *American Psychologist,* 1969, *24* (4), 409-429.

Campbell, D. T., & Stanley, J. C. *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally, 1966.

Clark, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior,* 1973, *12,* 335-359.

Clark, R. E. Constructing a taxonomy of media attributes for research purposes. *AV Communication Review,* 1975, *23*(2), 197-215.

Cohen, J. Multiple regression as a general data-analytic system. *Psychological Bulletin,* 1968, *10*(6), 426-443.

Cronbach, L. J., & Snow, R. E. *Aptitudes and instructional methods: A handbook for research on interactions.* New York: Irvington Press, in press. (1976)

Ebel, R. L. Some limitations of basic research in education. *Phi Delta Kappan,* 1967, 81-84.

Elman, A., Calfee, R. C., & Filby, N. N. Fractional design: An overview. Unpublished paper, Stanford University, School of Education, undated.

Filep, R. & Schramm, W. A study of the impact of research on the utilization of media for educational purposes: NDEA Title VII 1958-1968. Final Report of USOE study. ERIC ED 042 064, July 1970.

Fisher, R. A. *The design of experiments.* London: Oliver & Boyd, 1935.

Hall, D. C. Research methodology of NDEA Title VII-A: A descriptive and critical analysis. *AV Communication Review,* 1972, *20*(2), 117-134.

Heckman, R. W. *Aptitude-treatment interactions in learning from printed instruction: A correlation study.* Unpublished doctoral dissertation (University Microfilms 67-10, 202), Lafayette: Purdue University, 1967.

Issac, S., & Michael, W. B. *Handbook in research and evaluation.* San Diego: Robert R. Knapp, 1972.

Pereboom, A. C. Some fundamental problems in experimental psychology: An overview. *Psychological Reports,* 1971 *28,* 439 455.

Riecken, H. W., & Boruch, R. F. *Social experimentation: A method for planning and evaluating social intervention.* New York: Academic Press, 1974.

Salomon, G. Can we affect cognitive skills through visual media? Jerusalem: The Hebrew University, 1971.

Salomon, G. *Sesame Street in Israel: Its instructional and psychological effects on children.* New York: Children's Television Workshop, 1974. (Publication date, 1976.)

Salomon, G., & Clark, R. E. Re-examining the methodology of research on media and technology in education. *Review of Educational Research,* in press, scheduled 1976, *46* (4). (Also available from ERIC, ED 098 971.)

Samuels, S. J. Effects of pictures on learning to read, comprehension and attitudes. *Review of Educational Research,* 1970, *40,* 397-409.

Samuels, S. J., Biesbrock, E., & Terry, P. A. The effects of pictures on children's attitudes toward presented stories. *The Journal of Educational Research,* 1974, *67,* 243-246.

Snow, R. E. Representative and quasi-representative designs for research on teaching. *Review of Educational Research,* 1974, *44,* 265-292.

Shulman, L. S. Reconstruction of educational research. *Review of Educational Research,* 1970, *40*(3), 371-396.

Thoresen, C. E. The intensive design: An intimate approach to research. *The Counseling Psychologist,* in press.

Thoresen, C. E., & Elashoff, J. D. Some comments on "an analysis of variance model for the intrasubject replication design." R & D Memorandum #111. Stanford University, School of Education, 1973.

Van Dalen, D. B., & Meyer, W. J. *Understanding educational research.* (Rev. ed.) New York: McGraw-Hill, 1966.

Winer, B. J. *Statistical principles in experimental design.* (2nd ed.) McGraw-Hill, 1971.